Available Online: 15 July 2025 DOI: 10.54254/3049-7248/2025.24686

Study on influencing factors of education completion rate: research on the relationship between school construction and different levels of education based on linear regression

Chenbing Zhou

Department of Economics, University of Colorado Denver, Denver, USA

chenbing0402@outlook.com

Abstract. This study examines the relationship between government education expenditure (GEE) and enrollment rates across primary, secondary, and tertiary education levels, addressing a critical gap in empirical research on educational investment outcomes. Using cross-national data from UNESCO and the World Bank (1970-2023), we employ linear regression analysis with key control variables including compulsory education duration and qualified teacher ratios. Our findings demonstrate that GEE exhibits a statistically significant positive correlation with enrollment rates, particularly at primary and secondary levels. This effect is most pronounced in low-income countries, where each 1% increase in education expenditure as a percentage of GDP corresponds to a 2.15% rise in secondary enrollment rates. The analysis reveals that infrastructure investments - particularly in teacher quality and school facilities - serve as the primary mechanism through which GEE improves enrollment outcomes. At the tertiary level, the study identifies additional significant factors including internet access in educational institutions ($\beta=0.77$, $p<0.01$) and national income classification (50.61% higher enrollment in high-income countries). These results remain robust to alternative model specifications and heteroscedasticity adjustments. Our research contributes to the human capital and educational equity literature by quantifying the differential impacts of government spending across development contexts. The findings suggest that policymakers should prioritize: (1) targeted infrastructure investments in low-income regions, (2) teacher training programs, and (3) technology integration in schools. These interventions appear most effective for improving both enrollment quantities and educational quality. The study provides empirical support for SDG 4 (Quality Education) implementation strategies, particularly in developing economies where education gaps remain most persistent.

Keywords: education completion rate, linear regression, school construction, educational levels

1. Introduction

Government school construction, funded by education expenditure, directly impacts educational access and quality. Government education expenditure, through school construction, has significant impacts on educational outcomes.

This study fills this gap by exploring the relationship between government school construction, mainly measured by government education expenditure, and enrollment rates at different educational stages. Additionally, Other independent variables (e.g., compulsory education duration, qualified teacher proportion) are included to identify key drivers of enrollment rates. By pinpointing key drivers of enrollment rates, resources can be directed to high-yield sectors. This targeted approach maximizes the return on educational investment, effectively boosting enrollment and aligning with sustainable education development goals.

2. Literature review

The research by Max Roser and Esteban Ortiz Ospina [1] discusses education financing methods and global expenditure trends, but lacks analysis on expenditure-enrollment rate links.

Additionally, some studies have explored the impact of government compulsory education duration on enrollment rates in small-scale settings. Luis Diaz Serrano et al. analyzed the impact of primary compulsory education duration on secondary enrollment in non-OECD countries [2]. The study shows that in low-income countries, extending compulsory education has high economic value for children, but parents have low willingness to let their children continue their education. This study uses more data to explore not only low-income countries but also the impact of compulsory education years on enrollment rates in high-income countries, providing a broader and more comprehensive reference for studying the impact of government construction on enrollment rates.

Some studies show that in some developed countries, parents' income is positively correlated with their children's educational attainment. However, Katrine V. Løken's Family Income and Children's Education: A Natural Experiment of the Norwegian Oil Boom* [3] presents a relatively unique perspective. By analyzing a unique dataset of Norwegians born between 1967 and 1969, which measures the permanent family income of their children during adolescence, the study examines the long-term impact of family income on children's educational level. However, the context of this study is the Norwegian oil crisis in the 1970s, which meant a general increase in income. The study found no clear causal relationship between family income and children's educational level in Norway. In the study by Stephens et al. [4], numerous factors including compulsory education policies and laws are effectively explained as related to educational attainment. However, the data collection in this study is limited in the white population in U.S. states, which may impose certain limitations on the research. Our study will draw on broader data to explore the relationship between income level and educational level, to assess the impact of government investment in school construction on enrollment rates and obtain more general conclusions.

Olubunmi Kayode Aya Nwoye's article mainly discusses the impact of teacher development on educational attainment in Nigeria [5]. The article points out that improving the quality of the teaching staff is equivalent to improving student performance. However, this study does not mention the relationship between the proportion of qualified teachers at different educational stages and enrollment rates. In our study, we will further investigate this issue and indicate the impact of the proportion of qualified teachers on government school construction, further exploring how this variable affects enrollment rates.

In Stephens' study [4], several factors including compulsory education policies and laws are effectively explained as related to educational attainment. However, the data collection in this study is limited and more prefer to the white population in U.S. states, which may impose certain limitations on the research.

Some studies have also analyzed the proportion of schools with internet access for teaching purposes and shown significant differences across regions, but they do not analyze their specific relationship with educational outcomes. Benchea's study specifically examines the dynamics of internet use in Romania and its impact on students' academic engagement [6], concluding that there is a negative correlation between academic engagement and internet use. However, this article does not specifically focus on the digital divide between regions and its impact on educational attainment, in order to better determine the impact of government investment in school construction on enrollment rates.

3. Theoretical analysis

3.1. Human capital theory

The human capital theory posits that education and training are investments in human resources, which enhance individual productivity and economic value by improving skills and knowledge. In this paper, the impact of education expenditure and compulsory education duration on enrollment rates can be explained by the human capital theory. Government education investment and extending compulsory education duration will improve individuals' educational levels, thereby enhancing the quality of the labor force in the labor market and promoting economic development. According to the human capital theory, more education investment helps improve individuals' professional skills, knowledge reserves, and innovative capabilities, ultimately increasing their employment opportunities and income levels. This theory emphasizes the importance of education for economic development and social progress and supports policies to extend compulsory education years and increase education expenditure, arguing that these measures can effectively enhance a nation's overall competitiveness.

3.2. Education equality theory

The education equality theory emphasizes that fairness in educational opportunities is an important embodiment of social justice. The theory holds that everyone, regardless of social background or economic conditions, should have equal access to education. In this paper, extending compulsory education years and increasing education expenditure help promote educational equity, ensuring that children from different social strata and regions can receive basic education. This theory supports the government in bridging gaps in educational resources through policy measures, especially in deprived areas and among vulnerable groups. According to the education equality theory, education should not be attachment of wealth and social status, and all citizens should have the opportunity to achieve social mobility through education. This theory helps explain why government education expenditure and the extension of compulsory education years can effectively improve enrollment rates, particularly for low-income families and students in remote areas, as providing more educational opportunities can break the intergenerational transmission of poverty.

4. Empirical analysis

4.1. Data source

Our study uses cross-sectional data from the UNESCO Institute for Statistics and the World Bank. These data include information on government education expenditure, compulsory education years, school enrollment rates, household income levels, etc., from 1970 to 2023 [7, 8].

The data used in this study come from two authoritative sources: the UNESCO Institute for Statistics and the World Bank. These cross-sectional datasets provide comprehensive information on key educational and socioeconomic indicators, including government education expenditure, compulsory education duration, enrollment rates, and household income levels. Spanning a wide time range from 1970 to 2023, these datasets offer valuable insights into studying long-term trends and dynamic interactions between education investment and socioeconomic factors. The inclusion of macroeconomic indicators and individual education statistics allows for a detailed examination of the relationships between government policies, household income, and educational outcomes across different national contexts.

4.2. Model

In this paper, regression models are constructed for the gross enrollment rates at different educational levels. The proportion of government education expenditure in GDP (Gross Domestic Product) is selected as the main explanatory variable x_1, while other independent variables x_2-x_5 including compulsory education duration, national income classification, proportion of qualified teachers, and the proportion of schools with internet access for teaching purposes are used to jointly explain the dependent variables enrollment rate at different level y_1, y_2, and y_3. The meanings of variables in the model are shown in Table 1.

This model explores the relationship between gross enrollment rates at different levels and school construction primarily measured by government education expenditure. National investment in education is a crucial component of school construction. Other variables, including compulsory education duration, national income classification, proportion of qualified teachers, and the proportion of schools with internet access for teaching purposes, all influence gross enrollment rates and measure school construction.

Table 1. Descriptive variables

Variable name	Label
y_1 prienroll	the proportion of students in a country or region who receive education in the age range of enrollment (usually 6-12 years old)
y_2 secenroll	the proportion of students in middle and high school education.
y_3 terenroll	the proportion of students' participation at the university level and above.
x_1 spen	Public spending on education as a share of GDP (%), includes expenditures from various levels of government on school construction, teacher salaries, educational resources, and curriculum development.
x_2 eduyear	Duration of compulsory education (year), the minimum mandatory education period set by the government.
x_3 highincome	Dummy variable, =1 if it is high-income or upper-middle-income country, =0 if it is low-income or lower-middle income-country
x_{4i} teacher	Percentage of qualified teacher
x_{5i} internet	Proportion of schools with access to the internet for pedagogical purpose

Table 2. Descriptive statistics

VARIABLES	N	mean	sd	min	max
y1_prienroll	300	102.6	9.277	66.97	137.6
y2_secenroll	281	92.19	19.72	24.12	137.0
y3_terenroll	253	47.16	27.39	3.655	116.2
x1_spen	307	4.279	1.551	0.864	13.48
x2_eduyear	307	9.938	2.721	0	15
x3_highincome	307	0.720	0.450	0	1

Table 2. Continued

x4i_teacher	307	93.77	13.78	1.329	100
x5i_internet	307	73.42	34.69	0	100
x4ii_teacher	307	4.529	0.281	0.845	4.615
x5ii_internet	307	4.008	1.092	0	4.615
y1i_lgprienroll	300	4.636	0.0880	4.219	4.932
y2i_lgsecenroll	281	4.506	0.259	3.224	4.927
y3i_lgterenroll	253	3.659	0.730	1.538	4.764

4.3. Descriptive statistics

Detailed information on variables is in Table 1. Table 2 shows the statistical data of each variable. The following are specific explanations on variables' distributions.

Primary School Enrollment Rate (y_1): The proportion of school-age children (6–12 years old) enrolled in education averages 102.6%, slightly exceeding 100%—likely due to repeated counting from population mobility. A small standard deviation indicates minimal cross-country variation, reflecting sound education policies and resource allocation.

Secondary School Enrollment Rate (y_2): Enrollment for junior and senior high school age groups averages 92.19%, lower than primary education. A larger standard deviation signals significant cross-country disparities, highlighting uneven expansion of education systems and policy effectiveness.

Higher Education Enrollment Rate (y_3): Participation in university and above-level education averages 47.16%, far lower than secondary education. Extreme gaps between minimum and maximum values underscore stark regional disparities and insufficient emphasis on higher education.

Proportion of Education Expenditure (x_1): Government education expenditure as a percentage of GDP averages 4.279%. Despite a small standard deviation, large extreme value differences exist; increased expenditure is linked to improved education quality and enrollment.

Compulsory Education Duration (x_2): The mean statutory minimum education years are approximately 10 (variance: 2.721), indicating significant regional differences. Longer durations generally boost primary/secondary enrollment but raise costs.

Income Level (x_3): As a dummy variable classifying countries by income, 72% of the sample are high/upper-middle-income (variance: 0.450). Higher economic status correlates with greater investment in educational infrastructure and resources, driving higher enrollment.

Proportion of Qualified Teachers (x_4): Teachers with professional certificates average 93.77%, but wide max/min gaps show some countries have low ratios due to inadequate training or unequal resource distribution. This directly impacts teaching quality and students' progression to higher education.

Proportion of Schools with Internet Access (x_5): Schools using the internet for teaching average 73%, with large variance indicating infrastructure gaps in some countries. Internet access enhances enrollment by enriching resources and enabling distance education.

4.4. Discussion on functional forms

As Table 4 showing, Enrollment rates are comprehensively influenced by educational resource investment (education expenditure x_1, teacher qualification x_4, internet access x_5), policies (compulsory education duration x_2), and economic levels (income grouping x_3). Theoretically, these variables are positively correlated with enrollment rates, so a linear model is initially established. However, data characteristics require adjustments to the functional form.

The primary (y_1) and secondary (y_2) enrollment rates have high dispersion, while the higher education enrollment rate (y_3) is nearly uniformly distributed (Figures 7, 8, and 9). After logarithmic transformation, the model's explanatory power (R^2) for y_2 and y_3 significantly improves (Table 6), so natural logarithmic forms are taken for all three.

The standard deviations of teacher qualification rates (x_4) and internet access rates (x_5) are large and negatively skewed, but logarithmic transformation does not improve linearity (Figures 3, 4, 5, and 6), so original forms are retained.

4.5. Dummy variables

In the regression model, dummy variable x_3 is set to represent national income (Table 1). When $x_3 = 0$, it represents low-income and lower-middle-income countries; when $x_3 = 1$, it represents upper-middle-income and high-income countries. Instead of treating x_3 as a continuous variable, we convert non-numerical categorical data into numerical values through a dummy variable based on the database's national income classification. This intuitively demonstrates the impact of national income differences on enrollment rates across educational levels and facilitates the study of interactions between categorical and continuous variables.

Based on the discussion of function form, we use Ordinary Least Squares (OLS) to establish three multiple linear regression models (Equation 1-3):

$$ln(y_1) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + e_1 \quad (1)$$

$$ln(y_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + e_2 \quad (2)$$

$$ln(y_3) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4 + \gamma_5 x_5 + e_3 \quad (3)$$

In these parameters, α_0, β_0, and γ_0 are intercept terms, representing the enrollment rates at different levels when there is no influence from independent variables in the model. α_{1-5}, β_{1-5}, and γ_{1-5} are slope parameters, while e_{1-3} are error terms. The independent variables in this model have covered the main influencing factors related to the research question. Moreover, in the current research context focusing on the enrollment rates at specific educational levels, it can explain part of the variations in enrollment rates. Meanwhile, other independent variables in the error term e_i are unrelated to the independent variables included in the model. Although they affect dependent variables, they do not cause bias in the coefficient estimation of the existing independent variables in the model. Therefore, we initially assumed that there are no omitted variables in this model.

4.6. Multicollinearity

As shown in Table 3, by analyzing the correlation between independent variables through the covariance matrix (Table 3), the following significant relationships are found (at the 1% level).

Table 3. The Relationship Between Variables x_{1-5}

	x1 spen	x2 edu~r	x3 hig~e	x4i te~r	x5i in~t
x1 spen	1				
x2 eduyear	-0.00380	1			
0.947					
x3 highinc~e	-0.2171*	0.2528*	1		
0.000100	0				
x4i teacher	-0.1546*	-0.0422	0.1483*	1	
0.00660	0.461	0.00930			
x5i internet	-0.1599*	0.1437*	0.6413*	0.2804*	1
0.00500	0.0117	0	0		

Government education expenditure (x_1) is negatively correlated with income level (x_3), suggesting higher reliance on private funding in wealthy nations, negatively correlated with proportion of qualified teachers (x_4), and school internet access rate (x_5), indicating that high - income countries may rely more on non - government educational investment (such as private investment).

National income level (x_3) is significantly positively correlated with compulsory education duration (x_2), proportion of qualified teachers (x_4), and school internet access rate (x5) ($r = 0.41 - 0.58$, $p < 0.01$), showing a high synergy between economic level and educational resource quality.

School internet access (x_5) is positively correlated with compulsory education duration (x_2) and proportion of qualified teachers (x_4), suggesting policy linkages between digital infrastructure and teacher quality.

Multicollinearity mainly stems from the mediating role of economic level (x_3), but the Variance Inflation Factors (VIF) are all < 5 (see Table 3 footnote), not reaching the severe multicollinearity threshold (VIF > 10), so all variables are retained.

4.7. Robust standard errors

Table 4 shows that Model 2's explanatory power is significantly higher than Model 1, highlighting the importance of logarithmic forms for secondary enrollment analysis.

Table 4. OLS regression results of models with y1, y2, y3 in logarithmic form

	y1i_lgprienroll	y2i_lgsecenroll	y3i_lgterenroll
x1_spen	-0.0006	0.0215**	0.0656**
	(0.0033)	(0.0086)	(0.0270)
x2_eduyear	-0.0011	0.0282***	0.0351**
	(0.0019)	(0.0042)	(0.0140)
x3_highincome	-0.0082	0.1281***	0.5061***
	(0.0153)	(0.0324)	(0.0995)
x4i_teacher	-0.0010**	0.0008	0.0041
	(0.0004)	(0.0009)	(0.0032)
x5i_internet	-0.0001	0.0031***	0.0077***
	(0.0002)	(0.0004)	(0.0013)
_cons	4.7504***	3.7381***	1.7201***
	(0.0442)	(0.0984)	(0.3268)
Adj R-squared	0.0161	0.4890	0.4190
N	300	281	253
Prob > F	0.0822	0.0000	0.0000
Prob > chi2	0.0000	0.0000	0.5687

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

After conducting the Breusch - Pagan Test, the Prob > chi2 value for Model 3 is 0.5687. This means that at the 5% significance level, we cannot reject the null hypothesis, i.e., there is no heteroscedasticity in this model, and the homoscedasticity assumption in the Gauss - Markov theorem holds. For Model 3, the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

Thus, in Model 3, the adjusted R^2 (R-squared) is 0.4190, meaning that 41.90% of the variation in the gross higher education enrollment rate can be explained by the government education expenditure as a proportion of GDP, compulsory education duration, national income level, qualified teachers, and the proportion of schools with internet access for teaching purposes. Meanwhile, the F - value is 0.0000, indicating the model is overall significant. In summary, the selected five factors play a significant role in explaining the gross higher education enrollment rate.

However, we find that the Prob > chi2 values for Models 1 and 2 are 0.0000 (Table 4), meaning the null hypothesis is rejected, and these models exhibit heteroscedasticity. Therefore, we recalculate the standard errors using robust standard errors while keeping the original regression coefficients unchanged, so that subsequent hypothesis testing and confidence interval estimation can account for heteroscedasticity, making the significance tests of coefficients more accurate. Robust regression results (Table 5) further validate the significance of x_1 and $x_5$in Models 2–3, with coefficients stable under heteroscedasticity adjustments.

Table 5. Robust regression results of models with y_1, y_2, y_3 in logarithmic form

	y1i_lgprienroll	y2i_lgsecenroll	y3i_lgterenroll
x1_spen	-0.0006	0.0215**	0.0656***
	(0.0033)	(0.0089)	(0.0214)
x2_eduyear	-0.0011	0.0282***	0.0351**
	(0.0027)	(0.0052)	(0.0138)
x3_highincome	-0.0082	0.1281***	0.5061***
	(0.0125)	(0.0257)	(0.1032)
x4i_teacher	-0.0010**	0.0008	0.0041
	(0.0004)	(0.0009)	(0.0034)
x5i_internet	-0.0001	0.0031***	0.0077***
	(0.0002)	(0.0005)	(0.0013)
_cons	4.7504***	3.7381***	1.7201***
	(0.0575)	(0.1074)	(0.3348)
R-squared	0.0325	0.4981	0.4305
N	300	281	253
Prob > F	0.1788	0.0000	0.0000

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In Model 1, the original adjusted R^2 is 0.0161. After robust regression, R^2 increases to 3.25%, almost double the pre - adjustment value. This means 3.25% of the variation in the gross primary education enrollment rate can be explained by the government education expenditure as a proportion of GDP, compulsory education duration, national income classification, qualified teachers, and the proportion of schools with internet access for teaching purposes. Meanwhile, Prob > F increases to 0.1788, indicating the entire model is still not significant.

In Model 2, the original adjusted R^2 is 0.4890. After robust regression, R^2 increases to 0.4981, meaning 49.81% of the variation in the gross secondary education enrollment rate can be explained by the government education expenditure as a proportion of GDP, compulsory education duration, national income level, qualified teachers, and the proportion of schools with internet access for teaching purposes. Meanwhile, the F - value is 0.0000, indicating the model is overall significant. In summary, the selected five factors play a key role in explaining the gross secondary education enrollment rate.

5. Interpretation of results

In Model 1, the proportion of qualified teachers (x_4) exhibits a significant negative impact on the gross primary education enrollment rate ($\beta = -0.0010$, $p < 0.05$), contradicting expectations. This may stem from data-specific anomalies, such as flawed statistical methods for qualified teacher counts or irrational allocation of teaching positions in certain regions. These issues could lead to diluted teacher resources and poor management, thereby inhibiting enrollment. Other independent variables (government education expenditure as a proportion of GDP, compulsory education duration, national income level, and school internet access rate) show negative but insignificant effects ($p > 0.05$), suggesting that primary education enrollment is influenced by unmodeled complexities like regional educational traditions, family emphasis on primary education, and population dynamics (e.g., birth rate variations affecting enrollment baselines).

In Model 2, government education expenditure (x_1), compulsory education duration (x_2), high-income country status ($x_3 =1$), and school internet access rate (x_5) significantly and positively impact gross secondary education enrollment ($p < 0.05$), aligning with theoretical expectations. Increased investment enhances school facilities and teaching quality, extended compulsory education expands access, high-income economies enable better resource allocation, and internet access enriches educational resources. The proportion of qualified teachers (x_4) shows an insignificant positive effect, possibly due to mismatches between teacher qualifications and actual teaching capabilities, or distractions from non-teaching duties.

In Model 3 for tertiary enrollment, variables x_1, x_2, $x_3=1$, x_5 significantly correlate with higher enrollment ($p < 0.05$), as government investment expands university scales, strengthens research capabilities, and enriches scholarship systems, while a solid compulsory education foundation fosters qualified student pipelines. The insignificant impact of qualified teacher proportion (x_4) may arise from tertiary education's emphasis on comprehensive academic environments, where individual teacher influence is diluted by factors like research collaboration, interdisciplinary integration, and shared academic resources, making single-variable effects less prominent.

6. Conclusions

This study focuses on constructing regression models to analyze the influencing factors of gross enrollment rates at different educational levels. Through the rigorous analysis of three models, it reveals the differences in the impacts of key factors and their internal logic. The government education expenditure as a proportion of GDP, compulsory education years, national income level, and school internet access status have significant positive promoting effects on secondary and university enrollment rates.

Based on these findings, policymakers are advised to optimize the structure of education expenditure, increasing investment in infrastructure and research platforms for secondary and higher education; dynamically extending the duration of compulsory education and strengthening its implementation; accelerating the coverage of internet facilities in secondary and higher education institutions; optimizing the allocation of qualified teachers; and drawing on the experiences of high-income countries to narrow regional educational gaps. These measures aim to enhance enrollment rates across educational stages and foster sustainable educational development.

Abnormal results such as the significant negative proportion of qualified teachers in the primary school enrollment rate model highlight the complexity of influencing factors in primary education, meaning future research needs to continue exploring potential factors such as population structure, fairness in educational resource allocation, and micro - teaching quality control to better improve this study.

References

[1] Roser M. and Ortiz-Ospina E. (2016) - "Education Spending",OurWorldinData, https://ourworldindata.org/financing-education

[2] Díaz Serrano, L., & Pérez Reynosa, J. H. (2018). Impact of Duration of Primary Education on Enrollment in Secondary Education: Panel Data Evidence from Developing Countries.

[3] Løken, K. V. (2010). Family income and children's education: Using the Norwegian oil boom as a natural experiment. Labour Economics, 17(1), 118-129.

[4] Stephens Jr, M., & Yang, D. Y. (2014). Compulsory education and the benefits of schooling. *American Economic Review*, *104*(6), 1777-1792

[5] Olubunmi Kayode Ayanwoye. (2023). Effects of Teachers' Professional Development on Students' Academic Achievement. Qeios. https://doi.org/10.32388/01sp81

[6] Benchea, L. (2021). Analysis on the relationship between internet use and students' academic engagement. *Journal of e-Learning and higher education*.
[7] UNESCO Institute for Statistics via UN SDG Indicators Database, UN Department of Economic and Social Affairs, https://unstats.un.org/sdgs/dataportal/database
[8] UNESCO Institute for Statistics (UIS), Education, 2024, https://uis.unesco.org/bdds
[9] World Bank's World Development Indicators (WDI)., https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators

Appendices

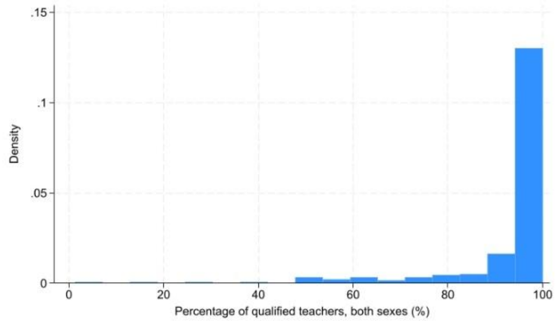
Appendix Table 1. The relationship between variables $x_{1\text{-}5,}$ with x_4 and x_5 in logarithmic form

	x1 spen	x2 edu~r	x3 hig~e	x4ii t~r	x5ii i~t
x1 spen	1				
x2 eduyear	-0.00380	1			
0.947					
x3 highinc~e	-0.2171*	0.2528*	1		
0.000100	0				
x4ii teacher	-0.0698	-0.0609	0.0787	1	
0.223	0.288	0.169			
x5ii inter~t	-0.1496*	0.3279*	0.5930*	0.1539*	1
0.00870	0	0	0.00690		

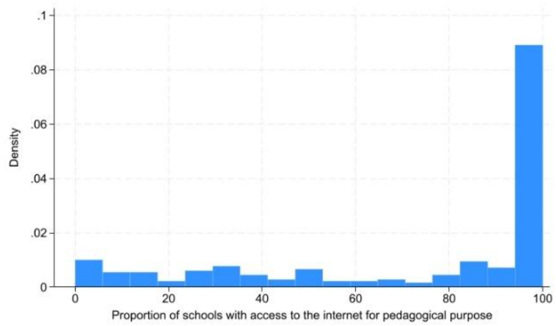
Appendix Table 2. OLS (Ordinary Least Squares) regression results comparison with x_4 and x_5 logarithmic form & models x_4 and x_5 are not in logarithmic form

	y1_prienroll	y2_secenroll	y3_terenroll	y1_prienroll	y2_secenroll	y3_terenroll
x1_spen	-0.0498	2.1154***	3.5160***	-0.1358	2.1144***	3.4915***
	(0.3528)	(0.6609)	(1.0701)	(0.3512)	(0.6609)	(1.0531)
x2_eduyear	-0.1328	1.3673***	-1.0424*	-0.1791	1.9384***	-0.3918
	(0.2086)	(0.3286)	(0.5747)	(0.2003)	(0.3204)	(0.5456)
x3_highincome	-1.6488	13.0834***	23.0266***	-1.1185	11.3319***	19.6350***
	(1.5011)	(2.3565)	(3.7144)	(1.6031)	(2.4953)	(3.8732)
x4ii_teacher	-2.4801	4.2585	7.4113			
	(1.9258)	(3.0533)	(5.1397)			
x5ii_internet	-0.3120	7.0016***	8.2997***			
	(0.6299)	(1.0025)	(1.6344)			
x4i_teacher				-0.1006**	0.1002	0.1958
				(0.0403)	(0.0678)	(0.1231)
x5i_internet				-0.0128	0.2172***	0.2809***
				(0.0208)	(0.0326)	(0.0518)
_cons	117.7558***	12.8774	-40.6424*	116.0968***	30.5141***	-16.9110
	(9.1426)	(14.4104)	(23.8947)	(4.6361)	(7.5690)	(12.7274)
Adj R-squared	0.0056	0.4779	0.3555	0.0246	0.4798	0.3745
N	300	281	253	300	281	253
Prob > F	0.2482	0.0000	0.0000	0.0304	0.0000	0.0000

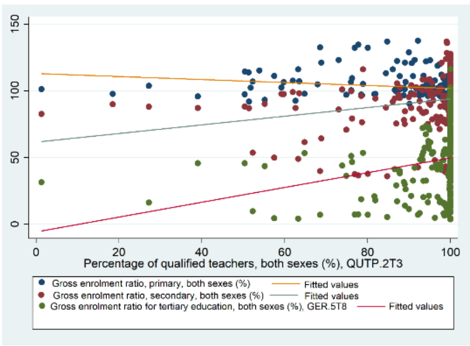
Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$



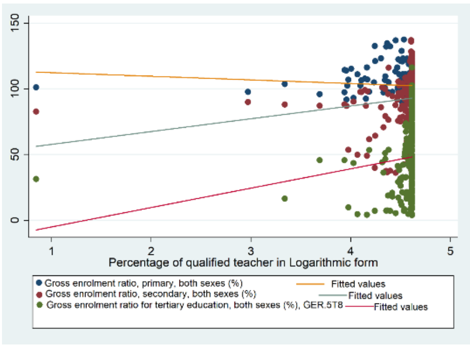
Appendix Figure 1. Histogram of percentage of qualified teachers



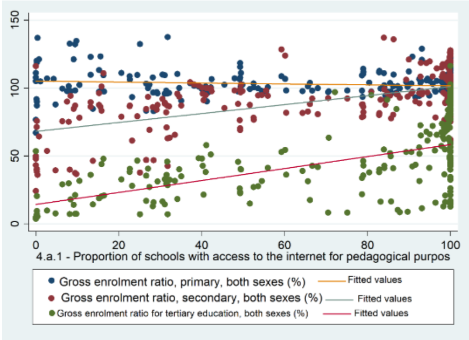
Appendix Figure 2. Histogram of proportion of schools with access to the internet teachers



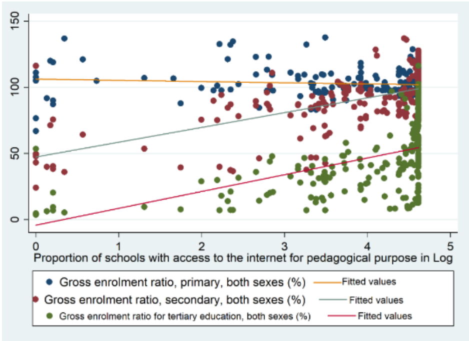
Appendix Figure 3. The relationship between gross enrollment ratio in different education levels and percentage of qualified



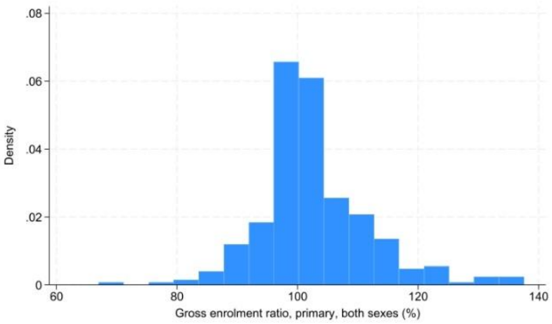
Appendix Figure 4. The relationship between gross enrollment ratio in different education levels and percentage of qualified teachers in logarithmic form



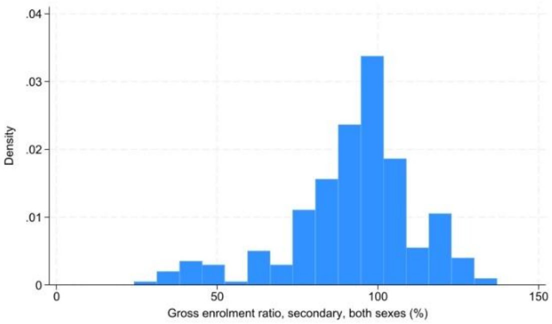
Appendix Figure 5. The relationship between gross enrollment ratio in different education levels and proportion of schools with access to the internet



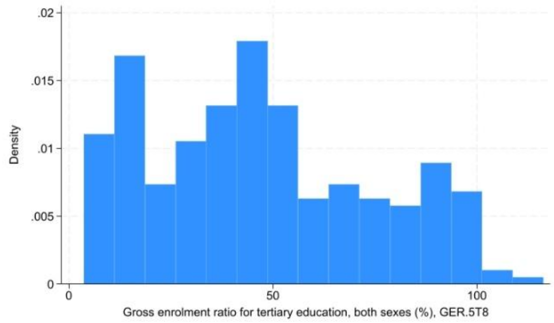
Appendix Figure 6. The relationship between gross enrollment ratio in different education levels and proportion of schools with access to the internet



Appendix Figure 7. Histogram of gross enrolment ratio for primary



Appendix Figure 8. Histogram of gross enrolment ratio for secondary



Appendix Figure 9. Histogram of gross enrolment ratio for tertiary